cTibTeX 0.6:  
Tibetan for \TeX and \LaTeX\n\textbullet\

Oliver Corff*  
corff@zedat.fu-berlin.de

July 1st, 2001

Abstract

cTibTeX is a package offering Tibetan support for \TeX and \LaTeX\n\textbullet\
This package is based on earlier works by Schwartz, Sparkes, Sirlin, 
Steiner and Preining (and would not be possible without their contri-
butions!) but in contrast to those the complete retransliteration pro-
cess is built on the ligature functionality of \TeX and Metafont, thus 
effectively eliminating the need for installing any external preprocessor
or \Omega mega.

Contents

1 Introduction 2
  1.1 Words of Thanks ............................................. 2
  1.2 Some Technical Notes ....................................... 3

2 Installation 4

3 User Commands 4
  3.1 Transliteration Table ....................................... 5
  3.2 Known Transliteration Problems ............................... 5
  3.3 Diacritical Marks ........................................... 5

4 Example 8

5 Legal Issues 12

6 Outlook and Desiderata 12

*This work was partially funded by DFG (Deutsche Forschungsgemeinschaft) in context with the Mongolian Lexicography Project.
1 Introduction

Support for Tibetan has been available for the \TeX and \LaTeX user community for quite a few years thanks to the contributions of Ronald Schwartz, Jeff Sparkes, Dominik Wujastyk\footnote{URL: \texttt{CTAN://tex-archive/language/tibetan/original/}}, Sam Sirlin\footnote{URL: \texttt{CTAN://tex-archive/language/tibetan/sirlin/}} and Beat Steiner\footnote{URL: \texttt{CTAN://tex-archive/language/tibetan/steiner/}}.

After studying these sources it became evident to me that these systems were composed mainly with a \TeX and \LaTeX user in mind who is willing to set up additional external software (the Latin–Tibetan converter/preprocessor) and who, in addition, is willing to cope with decidedly “non-\TeX-ish” code constructs (like \texttt{%%}) for marking Tibetan language portions. I considered it useful if external preprocessors and unidiomatic code could be avoided at reasonable cost for the user, i.e. without introducing loads of \texttt{\textbackslash ThisIsMySpecialTibetanLetter}-like commands. In addition, I appreciated the idea of a Tibetan support that seamlessly fits into the \TeX and \LaTeX world—after the typical \texttt{\input...} or \texttt{\usepackage{...}} declaration, the user should ideally write something like \texttt{\{\tib .bod skad.\}} in order to obtain \texttt{བོད་ཁ ི} which implies that the retransliteration engine is realized somewhere in the realms of \TeX and \LaTeX.

The astute observer may inject that Norbert Preining made Tibetan support without preprocessor available in 1997 for the \Omega mega system.\footnote{The \Omega mega home is: \texttt{http://www.ens.fr/omega}. \Omega mega is also shipped with recent versions of Thomas Esser’s \texttt{teTeX} (URL: \texttt{http://tug.org/teTeX/}), to name just one example.} As long as the transition to Unicode and \Omega mega has not been completed on a broad basis, I consider that there is still sufficient need for a \LaTeX 2ε-based Tibetan package. If its features were also usable in plain \TeX, I decided not to frown.

1.1 Words of Thanks

Of course, there is D. E. Knuth who created \TeX in the first place, and whose article on “The New Versions of \TeX and METAFONT”\footnote{Originally published in \textit{TUGboat} 10 (1989), 325–328; 11 (1990), 12; studied in Knuth: \textit{Digital Typography}, 1999, CSLI Publications, Stanford, California, p. 563–570.} ignited the initial spark for building a transliteration engine as a pure, if not huge ligature table. Ongoing discussions about Tibetan transliteration issues with Wolfgang Lipp in the context of the Pentaiglot Project were also very helpful. My special thanks go to Mr. Florian Reissinger for his patience while reviewing the glyph lists and his comments after proof-reading the test runs. The warm-hearted patience and understanding support by \texttt{བོད་ཁ ི}
and |\[\ddash\]| was essential for finishing this project.

1.2 Some Technical Notes

Readers who are not interested in the internals of cTib\TeX can safely proceed to sections 2 and 3 of this text. The following notes are intended for those who want to get an understanding of the workings behind the scenes.

In cTib\TeX, the Tibetan retransliteration system is based on the ligature mechanism located at the junction of Metafont and \TeX. In order to implement the ligature mechanism Sirlin's Metafont sources were rearranged completely. A separate coding mechanism with symbolic character codes was introduced so that all character interaction (numbering, ligtable programs etc.) could be defined on a purely mnemonic basis, thus avoiding the cumbersome inflexibility of a fixed numbering scheme. During this work a mistaken glyph was discovered which appeared twice: a second cwa (\[\ddash\]) showed up in the slot of gwa (\[\ddash\]) while the gwa glyph was missing altogether. A new gwa was then built using outline fragments of other glyphs.\(^6\)

Some character definitions were then taken from Norbert Preining's work; I added a text symbol visarga (\[\ddash\]) into the font.

In a further step, the vowel bounding boxes for [e i o] were raised while several duplicate [u]s with gradually sinking bounding boxes were added. Vowels could then be combined with consonants by simple ligature and kerning instructions, effectively eliminating the need for any \texttt{\textbackslash accent-}

construction on the document side. The exact alignment of carrier consonant and vowel can now be fine-tuned in the ligature table (\texttt{ctibligs.mf} in the \texttt{mfinput} directory). While most of the conventional Tibetan words can be written easily in cTib\TeX, the system, unlike Steiner's preprocessor, will fail with Sanskrit input as buddha which turns out as *\[\ddash\]\[\ddash\]** instead of \[\ddash\]**.

In cases like this users are requested to consult the vertical stacking command \texttt{\textbackslash V{}{}} which allows explicit stacking of any desired letter combination.

Finally, a style file was created (\texttt{ctib.sty}) which provides the user interface to the Tibetan script and related commands. The style file calls font definition files which were created for accommodating future extensions like new script styles and typefaces.

In early summer 2001, contact with Sam Sirlin was renewed, who had released the version 6.0 of his Tibetan package in the meantime. A major feature was the greatly enriched collection of glyphs (additional ligatures, letter elements for combining new and rare letters, diacritical marks, etc.) which was then made available to cTib\TeX.

\(^6\)Alas! The mistaken cwa which wanted to be taken for a gwa even infiltrated Omega Tibetan...
2 Installation

Installation of this software package is straightforward: The installation procedure depends on the nature of the actual \TeX{} system. The directory tree of e.g., \TeX{} is different from the emtex tree; hence the source archive ctib4tex.nn.zip features the following subdirectories the contents of which has to be placed into appropriate branches of the \TeX{} installation:

- **mfinput** holds the complete Metafont sources for the Tibetan fonts. The suggested path for emtex users is \emtex\mfinput\ctib; for \TeX{} users \$\TEXMF/fonts/source/public/ctib \$ is a suitable choice.

- **tfm** holds all necessary font metrics files. The suggested path for emtex users is \emtex\tfm\ctib; for \TeX{} users \$\TEXMF/fonts/tfm/public/ctib \$ is a suitable choice.

- **texinput** holds all style files, font encoding definitions etc. which are read by \TeX{} and \LaTeX{} \$2\epsilon$. The suggested path for emtex users is \emtex\texinput\ctib; for \TeX{} users \$\TEXMF/tex/latex/ctib \$ is a suitable choice.

- **doc** contains the documentation (the document which you are reading right now). It can be placed in \emtex\doc\ctib (for emtex users) or \$\TEXMF/doc/latex/ctib \$ (for \TeX{} users).

It may become necessary to rehash the directory database of the \TeX{} system. When in doubt, consult your system administrator or local \TeX{} wizard. On \TeX{} systems, the command texhash will perform this service.

3 User Commands

\LaTeX{} \$2\epsilon$ users activate Tibetan support for their documents via a \usepackage declaration

\documentclass[\textwidth,11pt]{article}
\usepackage{ctib}

in the preamble of the document, while \TeX{} users activate Tibetan support for their documents via an \input declaration

\input ctib

in the beginning of the document. The only command necessary is \tib which switches to Tibetan:

This is Tibetan:\
\[\texttt{\$\swasti.bod skad.$}\]

This is Tibetan:\$
\texttt{\$\swasti.bod skad.$}\$
The *tsheg* is generated automatically after every syllable and can be inhibited by the command `\note\tsheg`. So, `\tib go` creates \(\tilde{\eta}\) whereas `\tib go\note\tsheg` creates \(\tilde{\eta}\). The intersentence space after sentences ending in \(k, g\) is created by `\K` which removes the *tsheg* at the same time. Additional *tshegs* are produced by commata , while the full stop . generates the *shad*, the exclamation mark ! generates a *tsheg shad* and the colon : produces a *visarga*. *Swasti* can also be abbreviated as \(\oplus\). See also table 6 for a fairly complete overview of available special symbols.

Stacks of consonants used for expressing Sanskrit words are not necessarily contained in the basic glyph collection of cTib\TeX. They can, however, be generated easily with the `\V{}{}` command (V like *vertical*). An abbreviation exists also for *om* which is `\om`.

\begin{verbatim}
\v{\text{\tib \om, ma nxi pa}\V{}{\de\ma}\hrih:}\\
\v{\text{\tib \dme, ai.}}
\end{verbatim}

Abbreviations exist:

<table>
<thead>
<tr>
<th>Abbreviations exist:</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>\tib \dme \ai.</code></td>
</tr>
</tbody>
</table>

### 3.1 Transliteration Table

Unlike with Steiner's and Preining's systems, there is only one transliteration model available; at present the user has to accept what the system offers. Please consult the following tables for an overview of available symbols. Note: symbols which could be added with this version thanks to Sam Sirlin's recent work on Tibetan fonts were [framed] for easy identification.

### 3.2 Known Transliteration Problems

The transliteration services of cTib\TeX are coded in the ligature table of the font implying that these services know a lot about adjacent letters but know next to nothing about Tibetan syllables. The approach follows thus always a "first match" and not a "correct match" method. It is hence possible that consonant clusters are converted to Tibetan glyphs in a manner which was not intended by the writer. The hyphen - helps solve these ambiguities:

The syllable `brtsams` is \(\tilde{\text{\tib brtsams}}, \text{not}\ \tilde{\text{\tib b-rt\text{sams}!}}\)
<table>
<thead>
<tr>
<th>Letter</th>
<th>Transliteration</th>
<th>Tibetan</th>
</tr>
</thead>
<tbody>
<tr>
<td>ka</td>
<td>kha</td>
<td>ga</td>
</tr>
<tr>
<td>ts</td>
<td>tsa</td>
<td>dza</td>
</tr>
<tr>
<td>pa</td>
<td>pha</td>
<td>ba</td>
</tr>
<tr>
<td>zha</td>
<td>za</td>
<td>'a</td>
</tr>
<tr>
<td>ra</td>
<td>la</td>
<td>sha</td>
</tr>
<tr>
<td>ha</td>
<td>a</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: cTib\TeX\ in\ Transliteration\ and\ Tibetan:\ Alphabet

<table>
<thead>
<tr>
<th>Letter</th>
<th>Transliteration</th>
<th>Tibetan</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>i</td>
<td></td>
<td></td>
</tr>
<tr>
<td>u</td>
<td></td>
<td></td>
</tr>
<tr>
<td>e</td>
<td></td>
<td></td>
</tr>
<tr>
<td>o</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: cTib\TeX\ in\ Transliteration\ and\ Tibetan: Basic\ Vowels

<table>
<thead>
<tr>
<th>Letter</th>
<th>Transliteration</th>
<th>Tibetan</th>
</tr>
</thead>
<tbody>
<tr>
<td>tsa</td>
<td>thsa</td>
<td>dza</td>
</tr>
<tr>
<td>nxa</td>
<td>sha</td>
<td>kshxa</td>
</tr>
</tbody>
</table>

Table 3: cTib\TeX\ in\ Transliteration\ and\ Tibetan: Sanskrit\ letters
Seven basic consonants with aspiration \( h \) subjoined:

\[
\begin{array}{cccccc}
gha & jha & dha & bha & dzha & dxha & lha
\end{array}
\]

Nine basic consonants with \( y \) subjoined:

\[
\begin{array}{cccccc}
kya & khya & gya & pya & phya & mya & rya & hya
\end{array}
\]

Fifteen basic consonants with \( r \) subjoined:

\[
\begin{array}{cccccc}
kra & khra & gro & tra & thra & dra & pra
\end{array}
\]

\[
\begin{array}{cccc}
phra & bra & mra & dzra & zra & shra & sra & hra
\end{array}
\]

Six basic consonants with \( l \) subjoined:

\[
\begin{array}{cccc}
kla & gla & bla & rla & sla & zla
\end{array}
\]

Seventeen basic consonants with \( wazur \) subjoined:

\[
\begin{array}{cccccc}
kwa & khwa & gwa & cwa & chwa & nywa & twa & dwa & tswa
\end{array}
\]

\[
\begin{array}{cccc}
tshwa & zhwa & zwa & rwa & lwa & shwa & swa & hwa
\end{array}
\]

Twelve basic consonants with \( r \) surmounting them:

\[
\begin{array}{cccccc}
rka & rga & rnga & rja & rnya & rta
\end{array}
\]

\[
\begin{array}{cccccc}
rdra & rma & rba & rma & rtsa & rdza
\end{array}
\]

Ten basic consonants with \( l \) surmounting them:

\[
\begin{array}{cccccc}
lka & lgva & lnga & lca & lta & lla & lpa & lba & lha
\end{array}
\]

Eleven basic consonants with \( s \) surmounting them:

\[
\begin{array}{cccccc}
ska & sga & snga & snya & sta & sda & sna & spa & sha & sna & stsa
\end{array}
\]

Table 4: cTibTepX in Transliteration and Tibetan: Composites I
mand name, but for some of them there is also a short character mnemonic. Please consult table 6 on page 9.

4 Example

The following text, “The Story of Yug-pa-can the Brahman”, is blatantly stolen from Sirlin’s /doc/ directory. Some of the input conventions have, however, changed slightly, and so the full text in romanized source and Tibetan target forms is given here. Please note that the word and sentence spaces are not identical with those of Sirlin’s and Steiner’s systems, hence the line and paragraph layout is not completely identical. Modifications of input syntax are shown in the margins.

\swasti.
.yul zhig na bram ze dbyug pa can zhes bya zhig ’dug ste. rab du dbul ’phongs pa bza’ ba dang,. bgo med pa zhig go. des khyim bdag cig las ba glang zhig b-nyes te. nyin par spyad nas ba glang de khris de khyim bdag de’i khyim du song ba dang,. de na khyim bdag ni zan za ste. dbyug pa can gyis ba glang de khyim gyi nang du btang ba dang,. ba glang sgo gzhvan du song nas stor ro.. khyim bdag de zan de zos nas langs pa dang,. de na ba glang ma mthong nas des dbyug pa can la glang ga re zhes byas pa dang,. des smras pa. khyod kyi khyim du btang ngo,. khyod kyis nga’i glang bor gyis slar byin cig ces smras pa dang,. des smras pa. ngas ma bor ro.. de nas de gnyis ’grogs te. rgyal po’i thad du ’dong ba dang,. ’u ba cag gi rigs pa dang mi rigs pa rtog par ’gyur ro zhes smras nas de gnyis dong ba dang,. mi gzhvan zhig gi rta rgod ma zhig bros nas. des dbyug pa can la smras pa. rgod ma ma btang zhes smras pa dang,. des rdo zhig blangs te ’phangs pa dang rta’i rkang pa la phog nas rkang
<table>
<thead>
<tr>
<th>Generic Name</th>
<th>Alternative Symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>\tibvardfive</td>
<td>ꞕ</td>
</tr>
<tr>
<td>\tibvarsix</td>
<td>Ꞗ</td>
</tr>
<tr>
<td>\tibvarseven</td>
<td>ꞗ</td>
</tr>
<tr>
<td>\tibvareight</td>
<td>Ꞙ</td>
</tr>
<tr>
<td>\tibvarnine</td>
<td>ꞙ</td>
</tr>
<tr>
<td>\tibempty</td>
<td>Ꞛ</td>
</tr>
<tr>
<td>\tibShad</td>
<td>ꞛ</td>
</tr>
<tr>
<td>\tibTsheg</td>
<td>Ꞝ</td>
</tr>
<tr>
<td>\tibSwasti</td>
<td>ꞝ</td>
</tr>
<tr>
<td>\tibVisarga</td>
<td>Ꞟ</td>
</tr>
<tr>
<td>\tibTshegshad</td>
<td>ꞟ</td>
</tr>
<tr>
<td>\tibAlttshegshad</td>
<td>Ꞡ</td>
</tr>
<tr>
<td>\tibNyistshegshad</td>
<td>ꞡ</td>
</tr>
<tr>
<td>\tibChemgoshad</td>
<td>Ꞣ</td>
</tr>
<tr>
<td>\tibSbrulshad</td>
<td>ꞣ</td>
</tr>
<tr>
<td>\tibRgyagramshad</td>
<td>Ꞥ</td>
</tr>
<tr>
<td>\tibVarchemgoshad</td>
<td>ꞥ</td>
</tr>
<tr>
<td>\tibRjessungaro</td>
<td>Ꞧ</td>
</tr>
<tr>
<td>\tibSnaldan</td>
<td>ꞧ</td>
</tr>
<tr>
<td>\tibRnambcad</td>
<td>Ꞩ</td>
</tr>
<tr>
<td>\tibGtertsheg</td>
<td>ꞩ</td>
</tr>
<tr>
<td>\tibRinchenspungshad</td>
<td>Ɦ</td>
</tr>
<tr>
<td>\tibTopiniyigmgomdunma</td>
<td>Ɜ</td>
</tr>
<tr>
<td>\tibFinaliyigmgomdunma</td>
<td>Ɡ</td>
</tr>
<tr>
<td>\tibIniyiigmgomdunma</td>
<td>Ɬ</td>
</tr>
<tr>
<td>\tibAngkhyanggyon</td>
<td>Ɪ</td>
</tr>
<tr>
<td>\tibAngkhyanggyas</td>
<td>ꞯ</td>
</tr>
<tr>
<td>\tibHalanta</td>
<td>Ʞ</td>
</tr>
<tr>
<td>\tibLcirtags</td>
<td>Ʇ</td>
</tr>
<tr>
<td>\tibNyizlamaada</td>
<td>Ʝ</td>
</tr>
<tr>
<td>\tibHalf</td>
<td>Ꭓ</td>
</tr>
<tr>
<td>\tibirGteryigmgor</td>
<td>Ꞵ</td>
</tr>
<tr>
<td>\tibPaluta</td>
<td>ꞵ</td>
</tr>
</tbody>
</table>

Table 6: cTib\TeX Special Diacritics
pa beag go\K des smras pa. khyod kyis nga’i rta bsad kyis nga’i rta byin cig\K ci’i phyir rta sbyin. des smras pa tshur shog\K.
rgyal po’i drung du ’dong dang,. ’u bu cag gi zhal che gcod du ’long ngo zhes smras nas. de dag der song ba dang,. dbyug pa can des ’bras par b-rtsams te. des rtsig pa zhig gi steng nas mchongs pa dang,. de’i drung na thga ga pa zhig thags ’thag cing ’dug pa de’i steng du lhung nas thga ga pa de tshe ’phos pa dang,. thga pa’i chung mas dbyug pa can de bzung nas. khyod kyis nga’i khyo bsad kyis nga’i khyo byin zhig ces smras pa dang,. ngas khyod kyi khyo ci ltar sbyin zhes smras nas. tshur shog rgyal po’i drung du ’dong ngo,. des ’u bu cag gi zhal ce gcad do zhes dong ba las. lam gyi bar na chu bo gting zab po zhig yod de. chu de’i nang nas tshur shing mkhan zhig te’u kha na ’khier te ’long ngo,. de la dbyug pa can gyis chu’i gting ci tsam zhes dris pa dang,. chu’i gting zab bo zhes smras pas ste’u chur lhung ste. ste’u ma rnyer pa dang,. des dbyug pa can bzung nas. khyod kyis nga’i ste’u chur bskyur ro. des smras pa ngas ma bskyer ro. tshur shog rgyal po’i drung du ’dong dang,. des ’u bu cag gi zhal che gcad do zhes smras nas dong ngo,. . .continued...
eliminates the tsheg and inserts a long space.

...continued...
5 Legal Issues

cTib\TeX{} is published under the GNU Public Licence. And very much like I did when I picked up the work of my predecessors and changed a few things here and there, I warmly welcome to change and improve everything, but then: please rename the files. If not for the sake of source protection, then at least for making sure that various \TeX{} installations around the world do not get confused.

6 Outlook and Desiderata

At the time of this writing, cTib\TeX{} is actually already outdated as \Omega{}mega, the Unicode-capable \TeX{} successor, is available. Why, then, did I undergo this effort? I needed Tibetan now, for ongoing Mongolian lexicographical work which is all done using \LaTeX{} and Mon\TeX{}.

Though the font provided by Schwartz, Sparkes and Sirlin is already useful, I keep dreaming of a true Tibetan Metafont, not just outlines created by GNU fontutils. Such a genuine metafont would greatly facilitate the creation of new and different typefaces, at least with different weights, and the few Unicode characters not yet covered could then be produced easily.

Though this is a work slightly more targeted at \Omega{}mega than at \LaTeX{}\ 2\£, I am seriously considering preparing full-fledged Native Language Support for the standard document formats implying that all captions and the date and number formats have to be translated into Tibetan. With transliteration services provided internally, it should also be possible to integrate Tibetan into the Babel system.

Anyway, whatever the mistakes and the shortcomings are that have now crept into this Tibetan system, I can only kindly ask you to blame me, not my predecessors any more.